

平成21年6月9日現在

研究種目：基盤研究(B)  
 研究期間：2006～2008  
 課題番号：18300055  
 研究課題名（和文）知識発見システム MUSASHI の開発：開発支援環境の構築及びマイニング機能の強化  
 研究課題名（英文） Development of Knowledge Discovery System “MSUASHI” : Construction of Development Environment and Enhancement of Mining Functions  
 研究代表者  
 羽室 行信 (HAMURO YUKINOBU)  
 関西学院大学・経営戦略研究科・准教授  
 研究者番号：90268235

研究成果の概要：本研究では、データマイニング基盤ソフトウェアKGMOD(MUSASHIの後継ソフトウェア)の開発を行いインターネット上にて公開した。KGMODの開発においては、マルチスレッドに対応するなど先進的な技術を取り入れ、CSVテキストの大規模データをより効率的に解析処理することを可能とした。KGMODはライブラリとして実装されておりスクリプト言語に容易に統合することができる。KGMODを利用したデータマイニングの応用研究においてもいくつかの成果をあげている。

## 交付額

(金額単位：円)

	直接経費	間接経費	合計
2006年度	2,000,000	600,000	2,600,000
2007年度	7,300,000	2,190,000	9,490,000
2008年度	3,800,000	1,140,000	4,940,000
年度			
年度			
総計	13,100,000	3,930,000	17,030,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：データマイニング，知識発見，MUSASHI，ソフトウェア開発  
アルゴリズム

## 1. 研究開始当初の背景

大規模データベースからの知識発見を目的としたデータマイニング研究は、コンピュータサイエンスの分野において技術的側面についての研究は活況を呈している一方で、それらの技術の産業界への普及という応用の観点から見ると十分な成果があがっているとは言

がたい。これは、大学で研究開発されたソフトウェア技術を統一的に産業界に提供する枠組みが無いことに一因している。そこで、我々がこれまでに開発を進めてきた知識発見システムMUSASHIを拡張し、その枠組みを提供することで、データマイニングに関する大学での優れた研究成果を広く一般に普及させることが可能になると考えている。

## 2. 研究の目的

本申請における研究目的は、我々がこれまでに開発を進めてきた知識発見システム MUSASHI を拡張・充実させることにある。そのために、(1) 知識発見の分野の研究者および実務家が開発してきた知識発見手法のコンピュータへの実装を支援する開発環境を構築し、(2) また、その開発環境を用いて、新たな知識発見手法を開発し、その有効性を検証するための事例研究を行う。成果物は期間内に論文としてまとめていくと同時に、ソフトウェアとして実装し、実企業での利用に耐えうるシステム (KGMOD) として無償で公開する。

## 3. 研究の方法

### (1) 開発環境の構築

データマイニングアルゴリズムの効率的実装を考慮し、オブジェクト指向言語 C++ の採用および先進のアルゴリズムライブラリ BOOST C++ ライブラリを利用して開発を進める。MUSASHI の大幅な改良を伴う為に名称を KGMOD と改める。KGMOD は単一機能に特化したデータ処理関数およびマイニング関数の集まりであり、CSV テキストの標準入出力を共通インターフェースとして備えている。これらの関数をスレッド単位で動作させ、名前付きパイプで接続することで、柔軟なアプリケーション開発を可能とする。そしてそれらの関数の呼び出しは、多様なスクリプト言語からの呼び出しを可能とする。以上の機能を実現するソフトウェア開発を実施する。

### (2) 新たな知識発見手法の開発および応用研究

二つのデータベースにおいて、いずれかのデータベースに相対的に多頻度なパターンを顕在パターン (Emerging Patterns) と呼ぶ。そして顕在パターンに基づいた分類モデルの構築手法は CAEP (Classification by Aggregating Emerging Patterns) と呼ばれ、高い分類精度を達成できることがいくつかの研究で報告されている。この CAEP を時系列パターンおよび

分類階層 (Taxonomy) を扱えるように拡張し、KGMOD の一つの関数として実装する。さらには、分類階層の自動構築機能 (Indexing 機能) を研究開発する。以上の新たな知識発見手法を用い、Web ログ分析などの応用事例研究を実施する。

## 4. 研究成果

### (1) 開発環境の構築

上述の方法に従い KGMOD の開発を行いインターネット上に公開した (<http://kgmod.jp/kgmod/>)。マルチスレッド化により MUSASHI と比較して大幅な計算効率の向上を達成することができた。また KGMOD ライブラリに登録された個々の関数を呼び出すコマンド (M コマンドと呼ぶ) を開発し公開した。しかしながら、マルチスレッドプログラミングに想定以上の時間を要し、多様なスクリプト言語への埋め込みについては未完ではあるが鋭意開発を継続している。KGMOD の内容については国内外の学会において報告した。

### (2) 新たな知識発見手法の開発および応用研究

顕在パターンに基づいた分類モデルの構築手法を apriori アルゴリズムをベースとして実装し、KGMOD の関数の一部として公開した。この応用研究として、車両犯罪の発生についての研究を行い興味深い結果を得ることができた。さらに時系列と分類階層の機能を追加実装し、Web ログ分析の応用研究を行った。この研究によりデータ解析コンペティション 2006 にて最優秀賞 (<http://www.isc.senshu-u.ac.jp/~thc0640/dac18.html>)、および日本オペレーションズリサーチ学会事例研究賞 (<http://www.orsj.or.jp/aboutUs/award4.htm>) を受賞し、高い評価を得た。また、分類階層の自動構築機能についての応用研究としては、消費者購買要因データを利用したブランド選好についての研究を実施した。この研究によりマーケティング分析コンテストにて優秀賞を受賞し (<http://www.sendenkaigi.com/mac/2008/index.h>

tml)高い評価を得た。さらには、KGMOD を用いて食卓・食材データについての分析を行った。この研究によりデータ解析コンペティション 2008 にて最優秀賞を受賞し (<http://www.isc.senshu-u.ac.jp/~thc0640/dac.html>)高い評価を得た。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 10 件)

- ① 高橋宣行, 瀧澤重志, 加藤直樹, 具源龍, 「賃貸用オフィスビルのエントランスホールに対する感性評価のCAEPを用いた分析」, 日本建築学会計画系論文集, 74(640), 2009, pp. 1403-1410, 査読有
- ② 中原孝信, 森田裕之, 羽室行信, 本橋信次, 「中古車オークションデータを用いた自動車リースの査定条件に関する分析」オペレーションズ・リサーチ, Vol. 54 No. 2, 2009, pp. 65-72, 査読有
- ③ 羽室行信, 中西正雄, 山本昭二, 「統合化顕在パターン判別モデルによるWebアクセスログデータの分析」オペレーションズ・リサーチ, Vol. 53 No. 2, 2008, pp. 75-84, 査読有
- ④ 瀧澤重志, 材木敦史, 加藤直樹, 具源龍, 「新橋に立地するオフィスビルの感性評価を考慮した賃料分析」日本建築学会計画系論文集, 73(627), 2008, pp. 1053-1059, 査読有
- ⑤ D. Avis, N. Katoh, M. Ohsaki, I. Streinu, S. Tanigawa, Enumerating Non-crossing Minimally Rigid Frameworks, *Graphs and Combinatorics*, 23(1), 2007, pp. 117-134, 査読有
- ⑥ K. Yada, E. Ip, N. Katoh, Is this brand ephemeral? A multivariate tree-based decision analysis of new product sustainability, *Decision Support Systems*, 44(1), 2007, pp. 223-234, 査読有
- ⑦ A. Takizawa, F. Kawaguchi, N. Katoh, K. Mori and K. Yoshida: Risk Discovery of Car-Related Crimes from Urban Spatial Attributes Using Emerging Patterns, *International Journal of Knowledge-based and Intelligent Engineering Systems (KES)*, 11(5), 2007, pp. 301-311. 査読有
- ⑧ 具源龍, 瀧澤重志, 加藤直樹, 「航空写真からの屋根種類判別と稜線認識に関する研究-明度分布の曲線近似と平面回帰による2段階判別法の構築-」日本建築学会環境系論文集, 第611号, 2007, pp. 99-105, 査読有
- ⑨ 柳室純, 具源龍, 瀧澤重志, 加藤直樹, 豊田宏, 藤原淳, 小田憲史, 「画像解析・画像認識による膜材料劣化状況の自動診断技術の開発」膜構造研究論文集2006, 2007, pp. 71-82, 査読有
- ⑩ 若野洋平, 瀧澤重志, 加藤直樹, 「民間分譲マンションのモデルルーム来場者アンケートからの購買者の予測と分析」, 日本建築学会環境系論文集, 第606号, 2006, pp. 81-88. 査読有

[学会発表] (計 13 件)

- ① 中原孝信, 森田裕之, 羽室行信, 米田知弘, 「ポテンシャルモデルを利用したメニュー推薦システムの提案」, 平成20年度データ解析コンペティション最終成果報告会, 2009/3/30, NTTデータ豊洲センタービル(東京).
- ② T. Nakahara, H. Morita, "Collaborative Filtering by Using a Graph Partitioning Method on Binary Data", Asia Pacific Conference on Information Management (APCIM2009), 2009/3/27, Beijing University, China.
- ③ Cheung, S., Hamuro, Y., Morita, H., Katoh, N., "KGMOD: Powerful Data Processing and Mining Tool for KDD", Asia Pacific Conference on Information Management (APCIM2009), 2009/3/27, Beijing University, China.
- ④ 瀧澤重志, 具源龍, 加藤直樹, 「街路の視覚的情報を考慮した京都市伏見区中心部のひたくりの空間分析」, 建築学会第31回情報・システム・利用・技術シンポジウム, 2008/12/4, 建築会館ホール(東京).

- ⑤ 中原孝信, 森田裕之, 「グラフ分割問題を利用したCD購買データに対する推薦システム」, 経営情報学会2008年度秋季全国研究発表大会, 2008/11/9, 東北大学.
- ⑥ 森田裕之, 羽室行信, 加藤直樹, 中元政一, 松田康之, 「大容量データに対する高速データマイニングシステム-KGMOD-」, 経営情報学会2008年度秋季全国研究発表大会, 2008/11/9, 東北大学.
- ⑦ 森田裕之, 中原孝信, 羽室行信, 原田博実, 片本一平, 本橋 信次, 米田知弘, 「残価に影響を与える傷パターンを利用したオートリース戦略に関する分析」, 平成18年度データ解析コンペティション成果報告会, 2008/3/27, NTTデータ豊洲センタービル(東京).
- ⑧ 羽室行信, 中西正雄, 山本昭二, 「Webアクセスログデータを用いたデータマイニング手法のCRMへの応用」, 日本マーケティングサイエンス学会, 2007/12/8, 成蹊大学.
- ⑨ A. Takizawa, K. Yoshida and N. Katoh: Applying Graph Mining to Discover Substructures of Room Layouts Which Affect the Rent of Apartments, The 2007 IEEE International Conference on Systems, Man and Cybernetics (SMC2007), 2007/10/7, Montreal, Canada.
- ⑩ Naoki Katoh and Shin-ichi Tanigawa, Enumerating Constrained Non-crossing Geometric Spanning Trees, 13th Annual International Conference Computing and Combinatorics, 2007/7/17, Alberta, Canada.
- ⑪ 羽室行信, 中西正雄, 山本昭二, 「Emerging Sequence Patternに基づくWebアクセスログデータからの知識発見」, 日本マーケティングサイエンス学会, 2007/6/16, 関西

学院大学.

- ⑫ 羽室行信, 中西正雄, 山本昭二, 「Emerging Sequence Patternに基づくWebアクセスログデータからの知識発見」, 平成18年度データ解析コンペティション成果報告会, 2007/3/26, NTTデータ豊洲センタービル(東京).
- ⑬ 瀧澤重志, 川口史恵, 加藤直樹, 森健治, 吉田和生, 「Emerging Patterns を用いた都市の車両犯罪の発生に関する分析」, FIT2006 第5回情報科学技術レターズ, 2006/9/5, 福岡大学.

[図書] (計1件)

- ① 加藤直樹, 羽室行信, 矢田勝俊『データマイニングとその応用』朝倉書店, 2008, 194ページ.

[その他]

ホームページ情報

- ①<http://kgmod.jp/kgmod>
- ②<http://www.isc.senshu-u.ac.jp/~thc0640/dac18.html>
- ③<http://www.orsj.or.jp/aboutUs/award4.htm>
- ④<http://www.sendenkaigi.com/mac/2008/index.html>
- ⑤<http://www.isc.senshu-u.ac.jp/~thc0640/dac.html>

## 6. 研究組織

### (1) 研究代表者

羽室 行信 (HAMURO YUKINOBU)

関西学院大学・経営戦略研究科・准教授

研究者番号：90268235

(2)研究分担者

加藤 直樹 (KATOH NAOKI)

京都大学・工学研究科・教授

研究者番号：40145826

森田 裕之 (MORITA HIROYUKI)

大阪府立大学・経済学部・准教授

研究者番号：80295732

(3)連携研究者